Grant No. 335

Title: The molecular biology of nitrogen fixing nodules in common legumes.

Institution: The University of Pennsylvania

Semi-annual report: January 15, 1967

1. During the past six months the laboratory has gained momentum in all directions, administrative and scientific, although the threat of demolition hangs over our head. We conducted hydrogenase studies and tested the new all-glass respirometer; we collaborated with the Urbana Laboratories on nitrogen-fixation/growth studies on guar seed using the nodule bacteria recovered from the India material of 1965; we retested all the hemoglobin samples preserved from past soybean and lespedeza harvests in preparation for amino acid analyses (there were no field funds during the summer 1966); we arranged with the Beckman Company to conduct a large number of amino acid analyses on the nodule hemoproteins; we maintained correspondence with two stations in India which have sought our cooperation in problems related to field nodulation; we have made plans for full field work in 1967 starting with lupine in the wintertime, carrying through a second trip to India ; and with as much local field work as our laboratory move will permit; we have made a literature survey to come abreast of progress in the protein and amino acid fields; we have continually sought a new laboratory site and have one space offered that is suitable; we have met with University officials and initiated negotiations with the University to finance renovation of space. These topics are elaborated in the report that follows.

2. Laboratory facility:

a. Notice was received In October that the West Philadelphia Redevelopment Authority, in the interests of urban renewal, had achieved condemnation of the city block which contains this laboratory. The tentative date of demolition was given as 1 April 1967. The laboratory has really outgrown the present 816 square feet here, and a site of 1,000 to 1,500 square feet is sought. The laboratory has been invited to move with the present landlord to a warehouse located at 42nd Street and Woodland Avenue, within the University environs, where adequate space is available. The support of the University in the extent of $15,000 for construction of the facility has been requested (see correspondence-attached). Numerous c nferences have been held but the request has not been acted upon. With the precedent established, it is supposed that if necessary the rent could continue as a NASA obligation, within the grant.

b. Equipment problems:

i) The model 17 RCA Whirlpool no-frost freezer has proven defective. Three thaws occurred, spontaneously, which destroyed the remnant of the India collection of nodules but did not apparently damage the nodule extracts. Despite visits from Whirlpool technicians, it has not been possible to ascertain the fault.

ii) The all-glass differential respirometer purchased from the Gilson Medical Electronics Company, six months ago, is giving erratic measurements in the quantitative study of hydrogen gas uptake by nodule tissues (q.v.) There are theoretical advantages to this instrument, which is a modified Barcroft, but there may be drawbacks in design which render it impractical for use with hydrogen, at least. (i) The attachment of the plastic readout may permit gas leakage; (ii) it is difficult to measure the volume (of gas) contained on either side of the manometer. The company will exchange this instrument for a Warburg if the difficulties continue.

c. capital equipment still needed:

i) high speed centrifuge (Spinco, or Sorvall under development)
ii) amino acid analyzer (Beckman or Technicon are first choices)
iii) cell breaker (Sorvall preferred)

3. Field work: There was no harvest of legume nodules last summer owing to lack of funds a year ago, but lupine collecting is planned to start late in January with new funds received in October. The lupines of the San Francisco Bay area develop during the winter months, January through March or April. The altitude lupines develop nodules in June-July.

4. Scientific studies:

a. Exercise in theory:

Protein, if it evolves progressively to greater functional efficiency, does so through changes in its structure. Its structure depends upon the structure of its contained amino acids. Amino acid replacements are known to occur in analagous proteins recovered from different individuals and species. It is generally accepted that the basis for the replacement of an amino acid is a change in the genetic nucleotide code-triplet. Such change calls forth a substitute amino acid to occupy a particular site in the polypeptide chain, willynilly. World-wide attention has focussed on the coding of amino acids in protein synthesis and mutation. With the publication of more and more sequential analyses, it is becoming possible now to examine the structure of the amino acids actually involved in mutation and, presumably, evolution. This information precedes our appreciating the effective changes wrought in protein structure and hence in protein function during evolution. I wish to direct attention to this area of thought, which seems to have been ignored so far. These paragraphs will indicate what can be done, just as a start, with the protein data becoming available. Admittedly at this time any analysis of structural progression with proteins is exploratory, based on partly hypothetical data.

Fitch and Margoliash in Science 155:279, 1967 (the January 20th issue) published a family tree for 20 cytochrome c proteins. They also derived, with the aid of a computer, a hypothetical ancestral mammalian cytochrome c. Their table 4 takes 17 amino acids from the latter and shows their variation among a dozen contemporary mammalian species. This is termed "descent". Although the numbering of the amino acids cited in that table fails (so far as I can see) to conform with Margoliash's previously published cytochrome numbering (cf. for instance the Atlas of Protein Sequence, compiled by Eck and Dayhoff, 1966) it is nonetheless fascinating to study the

specified mutations as variants of a common ancestor.

I have taken the individual data of table 4, as best I could interpret them, and I have collected them into a single flow-chart inorder to get an overall view of just how the mutating amino acids succeed each other, without regard to locus or species. Figure 1 is the result, and it shows a remarkable pattern, particularly of convergence toward isoleucine (W). However, it is not intended to imply that the flow shown was postulated by the authors, or actually occurred, or would occur; it is suggested that certain of the amino acids or code-triplets seem to be focal points for change or end points of change. And it is emphasized that this could be an illusion from too scanty data.

Figure 2 develops figure 1 further by presenting the structures of the amino acids which are involved in figure 1, thus emphasizing the structural changes that would have to be introduced into the proteins by the mutations. Up to this point, however, we have dealt with directional flow from a hypothetical ancestral protein. The hypothetical and the imaginative can be eliminated by just thinking of amino acids as replacing each other, or interchanging, at particular loci in the polypeptide chain, contemporaneously.

Doing just that, Watson (in Heme and Hemeproteins, Academic Press, 1966, p. 65) drew upon his work with Perutz and Kendrew (J. Mol. Biol. 13:669, 1965) to list the 12 heme-related amino acids of myoglobin (i.e. those amino acids within van del Waals radius of the heme) with the known variants for 6 of the 12 sites. Figure 3 lists those variants.

It is interesting, now, to compare the amino acid mutations identified above in myoglobins with the mutations indentified in cytochromes c: there seems to be some difference and some sameness, along with a preponderant involvement of the valine/isoleucine interchange (an interchange between two non-polar constituents).

Returning to the cytochrome paper, Fitch and Margoliash list in their table 1 twenty common amino acids and the number of mutations needed (within the code-triplet scheme) to interchange each for another. Of the 380 theoretically possible replacements, 32 replacements (i.e. 16 reversible pairs) require three mutations to accomplish. These are:

asp ⇌ met	phe ⇌ glN
asp ⇌ try	phe ⇌ lys
asN ⇌ try	glu ⇌ ilu
cys ⇌ glu	glN ⇌ ilu
cys ⇌ glN	his ⇌ met
cys ⇌ lys	his ⇌ try
cys ⇌ met	met ⇌ try
phe ⇌ glu	try ⇌ ilu

Twelve of the 20 amino acids are thus involved in a triple-mutation. That leaves 8 amino acids which can interchange with any other by means of only one or two mutations.

Among the 12, I see:

2 di-carboxylic acids and their amides (the two)
2 sulfur-containing acids (the two)
3 aromatic acids (the three)
1 branched aliphatic acid
1 imidazolyl acid
1 di-amino acid.

Is it my imagination to associate almost all of the structurally more complex amino acids with the more difficult mutational interchange? Or is there perchance something selective built in with the triplet-coding system that facilitates certain amino acid interchanges and that makes it harder for other amino acid interchanges (possibly those more disruptive to a protein) to occur? The only large polar acid escaping the list is arginine.

This is as far as I can develop these data. It does seem as though there were more to amino acid replacements than coding changes in the genetic material. There must be involved the availability of an amino acid, its stereochemical suitability, its reactive suitability, etc. Furthermore, if two mutations are required to go from one amino acid to another, both known, the question occurs: what happened to the intermediary form? The intermediary must not have been a lethal, unless it is postulated that two mutations occurred together, a suggestion the statisticians oppose. (Rather it is proposed that mutations occur over generations.) Suppose, as is hypothesized in figure 1, glutamic acid amide is replaced by methionine. According to the triplet-code, the highly polar glutamine has gone via leucine or via lysine to S-containing methionine. A polar amide, then, is replaced by a less polar di-NH_2, C-6 lysine or by a non-polar C-6 leucine with the bulky methyl side chain; and one of the latter would be replaced by a seemingly structurally unrelated methionine. What would be the effect on the cytochrome, and what would be the advantage, if any, in evolution? The foregoing is so strikingly different from any biosynthetically-feasible rearrangements that I am aware of.

My program is concerned with the analysis and the evolution of primitive heme proteins. Structural discussion like the above pertains to the entire field of heme proteins and might well be extended. The intriguing question is that of the randomness of the nucleotide base mutations. Examination of more of the sequential analyses will tell whether there is merit to the question and others like it. For instance, did the evolving genetic code and the evolving amino acids somehow safeguard the future of the evolving living systems, by placing limits? etc.

b. Amino acid analyses of the heme proteins:

Beckman Instruments Inc. has agreed to run the nodule hemeoprotein analyses on a consultant basis; our laboratory will supply the hydrolysates. The initial cost will be $175 per acid or base hydrolysate; both types are required of each protein to determine the full complement of amino acids. For a publishable protein analysis there should be three acid hydrolyses of increasing duration to take care of the heat-labile acids as well as those which form heat-stable bonds. The Company can run an average of 4 analyses per week, and on this basis the cost would be reduced, though no figure has been provided. Correspondence has been with David A. Gayner at Palo Alto, and I have also spoken with John Deutschlander of Palo Alta. A few thousand dollars invested in such consultation will give the start we need for deciding how extensively the hemoproteins should be examined to accomplish our research aims.

c. Hydrogenase activity in legume nodules:

Crushed, water extracted guar and lupine nodules demonstrate an active hydrogenase system, and in a hydrogen atmosphere bring about the reduction of such indicator dyes as methylene blue and methyl viologen. This is in contrast with the similarly prepared soybean nodules, but resembles the activity in lespedeza. As mentioned earlier, the differential respirometer is giving erratic measurements of the gas uptake, and we plan to switch to the classical Warburg apparatus when the work is resumed in the fall.

d. Agricultural applications of the research:

i) We have had correspondence with a tea plantation in Assam, India, where a legume crop, dal, is rotated with the tea, and where our help in nodulation has been requested. We plan to add soil and seed samples from this area to our field study conducted in conjunction with the Urbana Laboratories. The Plant Quarantine Division of the USDA (Mr. Z. A. Barker) has been entirely cooperative with our program of bringing in foreign soils for study; it is a pleasure to work with this office.

ii) The Urbana Laboratories, through Mr. William Taylor, has completed several growth studies on guar where guar seed was inoculated with bacteria recovered from the India nodules of 1965. Mr. Taylor obtained spectacular growth results, with some evidence for a food factor involved. We need fresh guar to continue the study, and accordingly we plan a second harvest in India sometime during 1967; funds are available for this.

Respectfully submitted,

Elizabeth Thorogood

mail address:
222 South 40th Street
Philadelphia, Pa. 19104

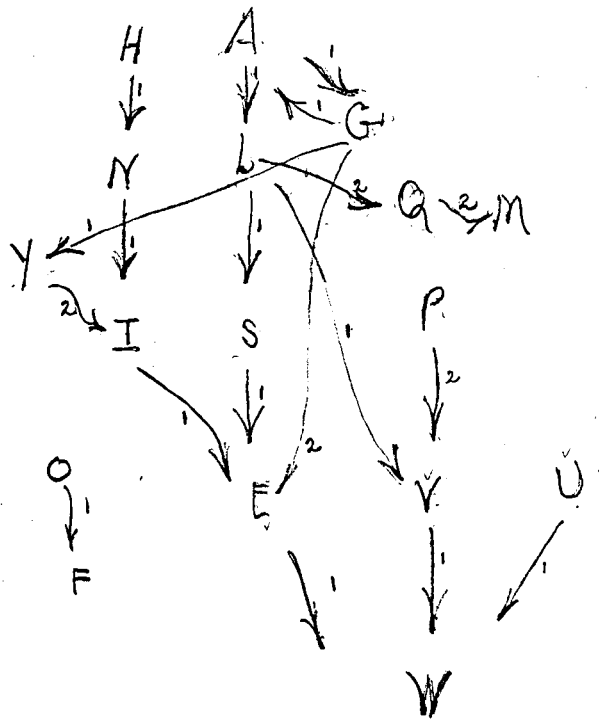
Elizabeth Thorogood, Ph.D.

FIGURE 1

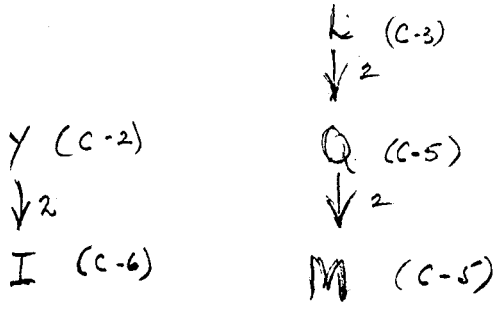
Seventeen (17) amino acids making 19 "descents" within 12 species of cytochrome C selected into one chart.

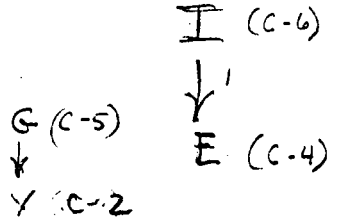
number is number of mutations required in the "descent" indicated.
1 = single mutation
2 = double mutation

A - asp
C - cys
E - glu
F - phe
G - gly
H - his
I - lys
L - ala
M - met
N - as-N
O - tyr
P - pro
Q - gl-N
R - arg
S - ser
T - try
U - leu
V - val
W - ile
Y - gly



(C-n) is length of C-chain





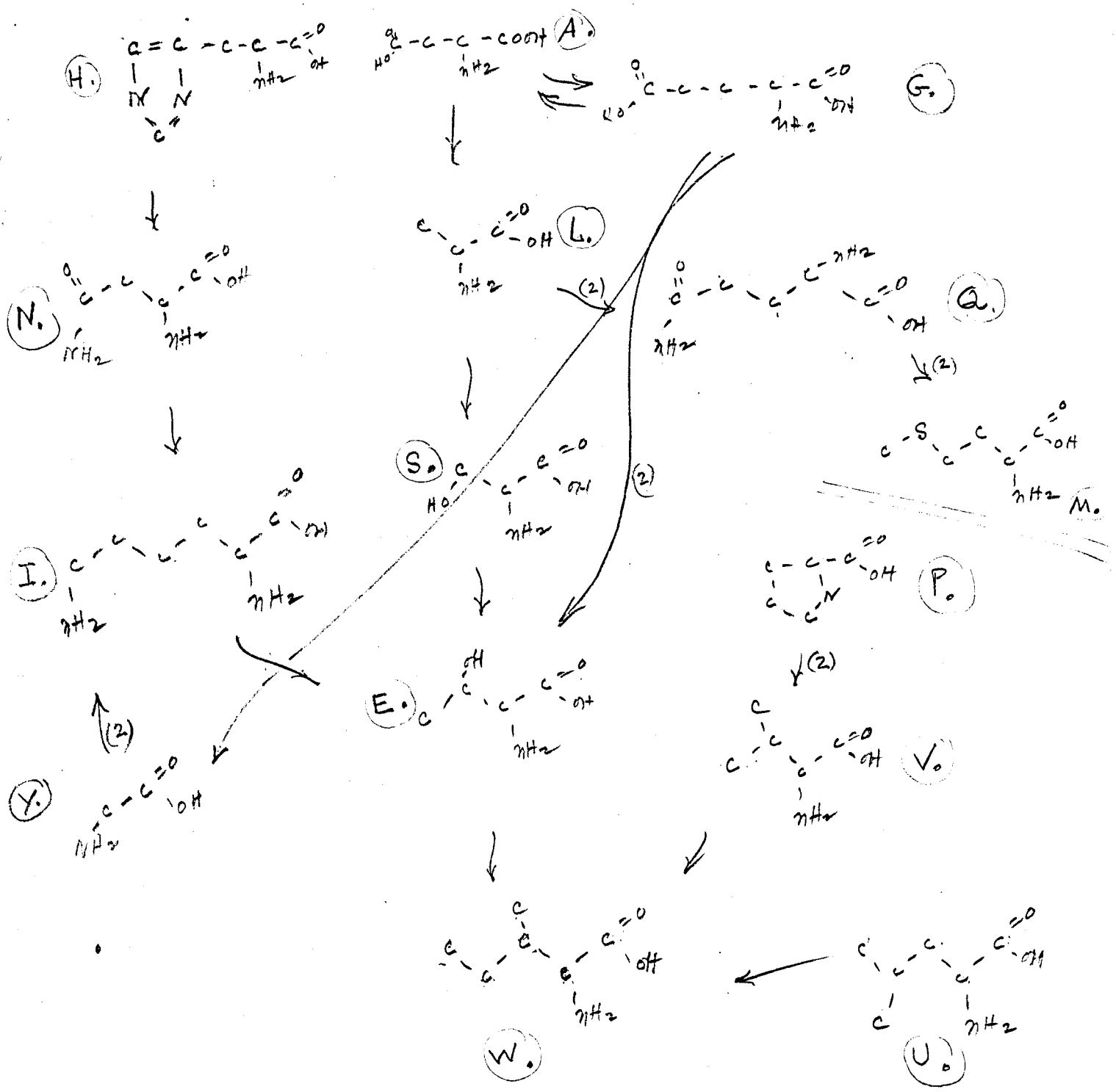


FIGURE 2a

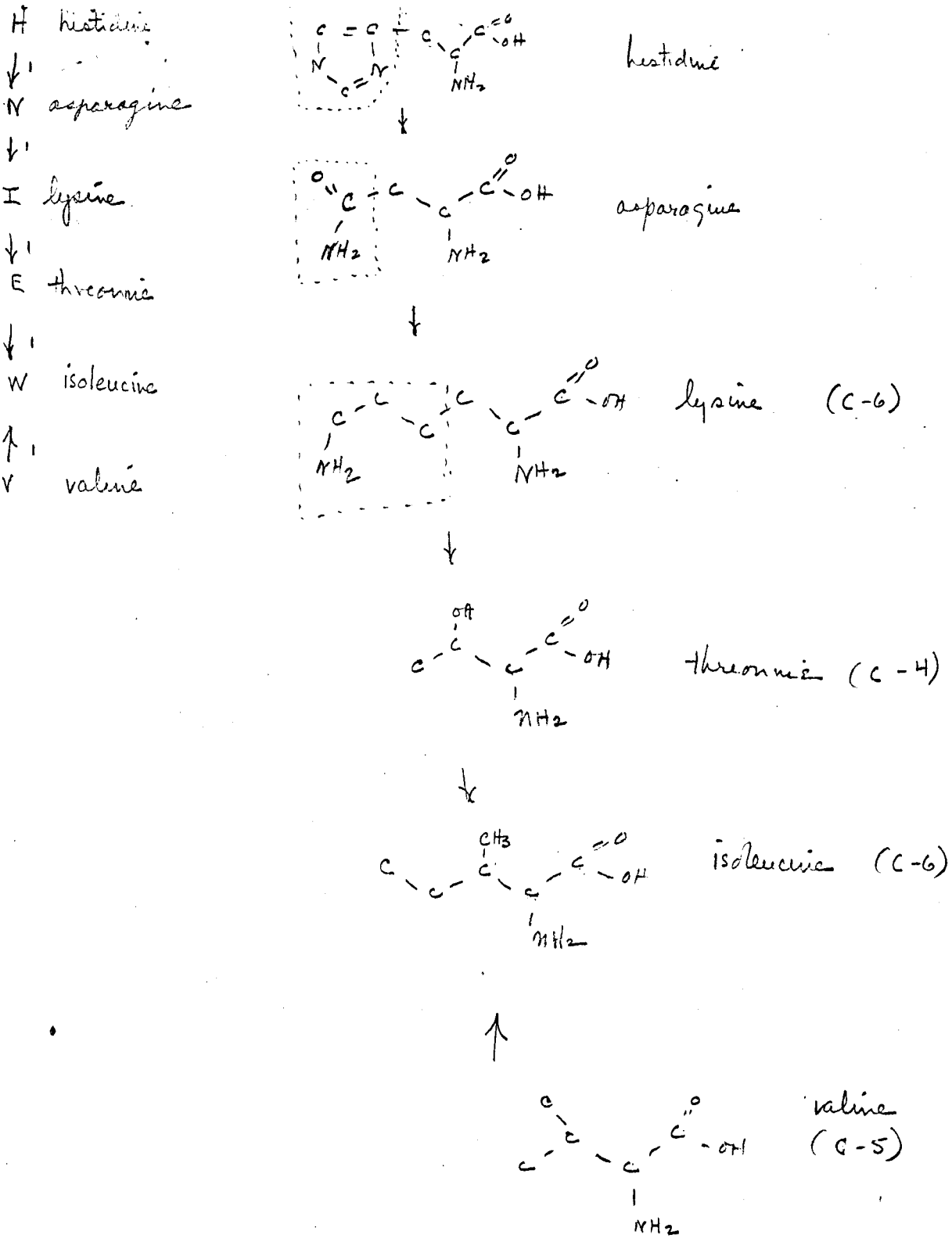


FIGURE 2 b

Section enlarged from Figure 2.

Replacements noted among the heme-related amino acids of myoglobin:

Y, L, S, A, and G

H and U

Y and W

Y, W, and U

Y and F

U and F

The increments denote associations seen in Figure 1, cytochrome C.

FIGURE 3.